

Academiejaar 2016-2017 – 1e examenperiode (Juni 2017)		Reeks 1
Opleidingsonderdeel: Onderzoekstechnieken - Oefeningen Campus: Schoonmeersen Lector(en): Anita Bernard, Jens Buysse, Bert Van Vreckem		Examendatum: 15/06/2017 10:30 Klas(sen): 2TI / 2TIN-TILE
Naam student:	Klas student:	
Geboortedatum student:	Stamnummer:	
Lector bij wie de student de onderwijsactiviteit volgde:	Lesgroep:	
Behaald resultaat: _____ op 21		

Tijdens het oefeningenexamen mogen volgende hulpmiddelen gebruikt worden:

- Eigen laptop: R, Excel (of andere rekenbladsoftware), rekenmachine, internetverbinding;
- Tabellen: z-tabel, t-tabel en χ^2 tabel;
- Afgedrukte cursus, eigen uitgewerkte oefeningen.

Onderlinge communicatie, op gelijk welke wijze, is uiteraard strikt verboden en leidt tot onmiddellijke uitsluiting.

Vul volgende vragen in. De voorziene ruimte is voldoende om te antwoorden. Antwoorden buiten de voorziene ruimte worden **niet** gequoteerd. Wanneer de uitkomst een reëel getal is, **rond dan af tot drie cijfers na de komma**.

Question	Points	Score
1	2	
2	4	
3	4	
4	6	
5	5	
Total:	21	

1. Beschouw het dataframe `mtcars` dat standaard in R zit.

(1pt) (a) Geef het gemiddelde van de variabele `mpg` (3 cijfers afgerond)

(a) _____

(1pt) (b) Geef Q_1, Q_2, Q_3

(b) _____

(b) _____

(b) _____

2. Een steekproef van 113 studenten levert een gemiddelde hartslag van 80,08 slagen/minuut. Men weet dat de standaard afwijking $\sigma = 12,13$ slagen/minuut.

(2pt) (a) Bereken het 95% betrouwbaarheidsinterval voor de gemiddelde hartslag in de populatie van de studenten en toon hoe je eraan komt. (rond je z-waarde af tot drie cijfers na de komma)

.....
.....
.....
.....

(2pt) (b) Bereken het 99% betrouwbaarheidsinterval voor de gemiddelde hartslag in de populatie van de studenten en toon hoe je eraan komt. (rond je z-waarde af tot drie cijfers na de komma)

.....
.....
.....
.....

3. Een uitvinder heeft een nieuwe, energie-efficiënte grasmaaimachine ontwikkeld. Hij beweert dat de motor gedurende 5 uur (300 minuten) voortdurend op een enkele liter gewone benzine loopt. Uit zijn voorraad van 2000 motoren selecteert de uitvinder een aselechte steekproef van 50 motoren. De motoren lopen gemiddeld 295 minuten, met een standaardafwijking van 20 minuten. Test de nulhypothese dat de gemiddelde looptijd 300 minuten is tegen de alternatieve hypothese dat de gemiddelde looptijd geen 300 minuten bedraagt. Gebruik $\alpha = 0,05$ en gebruik de methode met de p-waarde. (Veronderstel dat de looptijden van motoren normaal zijn verdeeld en rond tussentijdse resultaten af tot drie cijfers na de komma)

(1pt) (a) Geef H_0 en H_1

.....
.....
.....
.....

(2pt) (b) Voer de test uit

.....

.....

.....

.....

(1pt) (c) Wat concludeer je?

.....

.....

.....

.....

4. Open het bestand `breakingbad.csv`. We vinden hierin informatie rond de ratings ingegeven in IMDB voor de serie `breaking bad` (echte gegevens). We zijn geïnteresseerd in de volgende variabelen:

UserRating Gemiddelde user rating per aflevering

ID Identificatie van een aflevering

Season.episode Seizoen identificatie gevolgd door de identificatie van het afleveringsnummer van dat seizoen

(1pt) (a) Geef het meetniveau voor `UserRating`

(a) _____

(1pt) (b) Geef het meetniveau voor `Season.episode`

(b) _____

We zijn geïnteresseerd of er verband bestaat tussen de `ID` & `UserRating`.

(1pt) (a) Geef het wiskundig verband indien we een lineair verband wensen te onderzoeken (niet afronden).

.....

.....

.....

(1pt) (b) Geef de waarde van de metriek die de sterkte van dit verband geeft en interpreteer die.

(b) _____

.....

.....

.....

- (1pt) (c) Geef de waarde van de maat die informatie geeft over de mate waarin het opgestelde model de werkelijke data benadert en interpreteer deze maat.

(c) _____

.....

- (1pt) (d) Geef de gemiddelde scores voor seizoen 1 en seizoen 5. (grep1 en reguliere expressies kunnen je hierbij helpen - niet vergeten af te ronden)
 Seizoen 1:

(d) _____

Seizoen 5: :

(d) _____

5. Beschouw de studie gevoerd met als resultaat kids.csv. Er werd aan de kinderen gevraagd in verschillende schooldistricten wat ze belangrijk vinden. Ze gaven aan of dat goede punten, sportief zijn of populariteit het belangrijkste was voor hen. Ze moesten ook volgende elementen rangschikken: schoolresultaten, sport, uiterlijk en geld in de orde van belangrijkheid voor hen. Er werd ook gevraagd naar hun geslacht, hun graad en andere informatie. Volgende variabelen worden gedefinieerd:

Gender : Boy or girl

Grade : 4, 5 or 6

Age : Age in years

Race : White, Other

Urban/Rural : Rural, Suburban, or Urban school district

School : Brentwood Elementary, Brentwood Middle, Ridge, Sand, Eureka, Brown, Main, Portage, Westdale Middle

Goals : Student's choice in the personal goals question where options were 1 = Make Good Grades, 2 = Be Popular, 3 = Be Good in Sports

Grades : Rank of "make good grades"(1=most important for popularity, 4=least important)

Sports : Rank of "being good at sports"(1=most important for popularity, 4=least important)

Looks : Rank of "being handsome or pretty"(1=most important for popularity, 4=least important)

Money : Rank of "having lots of money"(1=most important for popularity, 4=least important)

- (1pt) (a) Geef de hypothesen indien we een hypothesetoets willen uitvoeren die het verband zoekt tussen de variabelen goals & grade

.....

- (2pt) (b) Voer de toets uit gebruik makende van de methode met kritieke grenswaarde evenals met de p-waarde (gebruik $\alpha = 0.01$).

.....

.....

.....

.....

.....

.....

- (1pt) (c) Geef de naam van de metriek en de metriek zelf om de sterkte van dit verband uit te drukken.

(c) _____

- (1pt) (d) Welke conclusie trek je hier uit?

.....

.....

.....